DOI: 10.48011/asba.v2i1.995

On an Alternative Susceptible-Infected-Removed Epidemic Model in Discrete-time ⋆

Jorge A. Costa Jr. * **Amanda C. Martinez** * **José C. Geromel** *

* *School of Electrical and Computer Engineering, UNICAMP, 13083-852, Campinas, SP, Brazil*

Abstract: This paper presents a new probabilistic dynamic model of the SIR class that describes, with appropriate precision, the temporal behavior of epidemics in discrete-time. Determination of the set of invariance and convergence conditions towards equilibrium are established. For numerical analysis, data of daily number of new diagnosed cases provided by the Brazilian Ministry of Health and World Health Organization of COVID-19 epidemic that currently occurs in Brazil is used. Illustrations and model prediction analysis are provided and discussed from full data of Italy, a country where the epidemic has already ended. The same ideas used on the development of the proposed model formulated in discrete-time may be adopted for continuous-time modelling as well. Three different and complementary strategies for parameter identification using the daily data available are considered.

Keywords: SIR epidemic models, Discrete-time systems, Least square fitting

1. INTRODUCTION

We are living in a new time characterized by an unprecedented demand of the health system. In Brazil, we have the Unified Health System – SUS – that is showing its importance to better serve the population of our country. It is necessary to highlight the commitment and dedication of all health care professionals and a significant part of the population, that also deserves praise, as they try to maintain effective social distancing, even in the face of unreasonable opinions that, against the majority of the world, insist on minimizing its beneficial effects.

From the public health point of view, it is necessary to be able of evaluating possible scenarios and validating actions in order to flatten the peak of the epidemic and preserve the hospitals service capacity.

Social distancing is perhaps the only action we have at the moment, but the key question is how to assess its effectiveness and how to decide when and how to mitigate it, without allowing a second epidemic wave. The research effort on mathematical models development appears to be a possible way to find an adequate answer, since a sufficiently precise model would be an appropriate device to predict the epidemic time evolution.

The literature presents countless studies dealing with epidemics. The deterministic modeling presented in the seminal paper Kermack and McKendrick (1927), almost a century ago, establishes a solid mathematical basis for continuous-time modeling. This result gave rise to the model classes known by the acronym SIR and more specific sub-classes, see the survey paper Hethcote (2000) for more details. Furthermore, it is important to mention the books Anderson (1987), Bailey (1975) and Brauer and Castillo-Chavez (2012) including the references therein as excellent sources of information on model development analysis, control design and many other related topics on epidemics. Discrete-time SIR classes modeling and applications can be found in Brauer and Castillo-Chavez (2012) and Anastassopoulou et al. (2020).

SIR models are expressed by nonlinear differential or difference equations, in continuous or discrete-time, respectively. The order depends on the number of state variables needed to discriminate the various classes of individuals in the population. Recently, in Giordano et al. (2020), a complete $8^{\underline{th}}$ order continuous-time model of class SIR has been proposed to evaluate possible scenarios of COVID-19 epidemic evolution in Italy. The paper Bertozzi et al. (2020) has established composed modeling useful for forecasting the infection spread in the population.

In general, models are parameter dependent. Naturally, they must be determined in such a way that the final model is as faithful as possible in the face of reality. As data on the COVID-19 outbreak is provided daily, it seems more natural to develop a SIR model in the discrete-time domain. In this framework, we present a new epidemiological, probabilistic, nonlinear, discrete-time varying model. Its invariant set is calculated and convergence towards equilibrium points is analysed. In our opinion, the proposed model is a valid theoretical alternative to the classical SIR model but its final validation needs to be established in practice. We note that, focusing on greater accuracy, it is necessary to allow its parameters to vary over time in order to capture trends in how the population behaves during the outbreak evolution. The proposed model is applied to

⋆ This research was in part supported by the Brazilian National Council for Scientific and Technological Development (CNPq) – grant numbers 302013/2019-9 and 141289/2018-0.

data of COVID-19 outbreaks that is presently occurring in Brazil and has already reached its end in Italy.

The notation used throughout the paper is standard. Specifically, the symbols $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{N}$ denote the sets of real, real nonnegative, and natural numbers, respectively.

2. CLASSIC AND PROPOSED SIR MODELS

The classic Susceptible-Infected-Removed (SIR) model in continuous-time and in the deterministic framework has been first proposed and analysed in the seminal paper Kermack and McKendrick (1927). Since then, many generalizations to deal with specific aspects of epidemics have been developed, as for instance, SEIR, SEIRS, SIRS, SEI, SEIS, SI, SIS, among others. The paper Hethcote (2000) provides an interesting overview on this matter from a unified viewpoint.

The set of natural numbers is $\mathbb{N} = \{0, 1, 2, \cdots\}$ and the independent variable $k \in \mathbb{N}$ defines time, measured in days. The initial time instant, $k = 0$, corresponds to the day on which the first case of infection was diagnosed [1]. The population $\mathcal{P}$ made up of M individuals [2] is split into three classes, and at each time instant $k \in \mathbb{N}$, each individual is supposed to belong to only one of them, namely:

- **Susceptible (S) -** is the group of healthy individuals. The total number of elements in this set, denoted by $s(k)$, indicates the number of healthy individuals on day $k \in \mathbb{N}$.
- **Infected (I) -** is the group of infected individuals. The total number of elements in this set, denoted by $i(k)$, indicates the number of infected individuals, capable of transmitting the disease, on day $k \in \mathbb{N}$.
- **Removed (R) -** is the group of individuals who no longer have the ability to transmit the disease because they are immunized or dead [3]. The total number of elements in this set is denoted by $r(k)$, with $k \in \mathbb{N}$.

By assumption, the population remains constant throughout the epidemic horizon, births are not taken into account, which implies $s(k) + i(k) + r(k) = M$ for all $k \in \mathbb{N}$. Let $x(k)$ be any member of the population on day $k \in \mathbb{N}$. The probability that he is healthy, infected or removed is $s(k)/M$, $i(k)/M$ or $r(k)/M$, respectively. Thus, the average number of individuals in each of these classes is $s(k)$, $i(k)$ or $r(k)$. The key issue of SIR models is the mechanism that determines the susceptible to infected transition which is responsible for its nonlinear nature. In the mathematical framework of probabilistic models, the ones of interest are developed and interpreted in the sequel.

2.1 The Classic SIR Model

Following Hethcote (2000), consider an arbitrary day $k \in \mathbb{N}$. Let $\beta(k)$ be the average number of effective contacts, those which result on infection of a person at time $k \in \mathbb{N}$. Under this assumption, $\beta(k)(i(k)/M)$ is the average number of effective contacts with infected of one individual of the population. Taking into account that at time $k \in \mathbb{N}$ the number of susceptibles is $s(k)$, then the average number of new infected is

$$n(k) = \beta(k) \left(\frac{s(k)}{M} \right) i(k) \tag{1}$$

In the aforementioned reference the time-invariant case $\beta(k) = \beta$ is considered. Moreover, by comparison with the model resulting from the mass action law, the parameter dependence $\beta = \eta M^{\upsilon}$ for some (η, υ) is discussed. Measurements strongly suggests that $\upsilon \approx 0$. This important aspect will be addressed in the next subsection.

2.2 The Proposed SIR Model

For an experiment, in a day $k \in \mathbb{N}$, let a pair of individuals (x_1, x_2) from the population M be randomly chosen, with replacement [4]. The probability that x_1 is healthy ($x_1 \in S$) and x_2 is infected ($x_2 \in I$), or vice versa, is $2(s(k)/M)(i(k)/M)$. Assuming that, with probability $p(k)$, a healthy person becomes infected whenever he meets an infected person, then the expected value of the number of new people infected is given by

$$\begin{aligned} n(k) &= s(k) \times p(k) \times 2 \left(\frac{s(k)}{M} \right) \left(\frac{i(k)}{M} \right) \\ &= \gamma(k) \left(\frac{s(k)}{M} \right)^2 i(k) \end{aligned} \tag{2}$$

where $\gamma(k) = 2p(k) \in (0, 2)$. Note that the first term $s(k)$ in the product shown in the first equality of (2), indicates that only healthy people, when meet infected people, can become infected. At this point it is interesting to compare the estimations for the number of new infected provided by both SIR models. From (1) and (2) it follows that

$$\beta(k) = \gamma(k) \left(\frac{s(k)}{M} \right) \tag{3}$$

from which some conclusions can be drawn. First, at the very beginning of the epidemic evolution the fact that $s(k) \approx M$ imposes $\beta(k) \approx \gamma(k)$ meaning that both models coincide. Of course, the same fact does not remain true whenever the epidemic evolves in time and the number of susceptible persons becomes smaller. Second, $\beta(k)$ may depend on many factors, as for instance the behavior, at least in part of the population, changes due to alerts and awareness campaigns.

Defining the index $\nu = \{1, 2\}$ such as $\nu = 1$ selects the classic SIR model studied in Anastassopoulou et al. (2020) and $\nu = 2$ selects the proposed SIR model, adopting the previous assumptions, the time evolution of the number of susceptible, infected and removed individuals in the population can be expressed in the unified form

$$s(k+1) = s(k) - \gamma(k) \left(\frac{s(k)}{M} \right)^{\nu} i(k) \tag{4}$$

$$i(k+1) = i(k) + \gamma(k) \left(\frac{s(k)}{M} \right)^{\nu} i(k) - \alpha(k) i(k) \tag{5}$$

$$r(k+1) = r(k) + \alpha(k) i(k) \tag{6}$$

with $\alpha(k) \in (0, 1)$, $\gamma(k) \in (0, 2)$ and nonnegative initial condition $s(0) = s_0$, $i(0) = i_0$ and $r(0) = r_0$ satisfying

[1] In Brazil, it occurred on February 26, 2020.

[2] The current population of Brazil is 210 million inhabitants, that is, $M = 210 \times 10^6$ individuals.

[3] We are not considering the possibility that people in the removed group may become susceptible again and be infected more often, as there is not enough evidence for this yet.

[4] Since $M \gg 1$ the pair of individuals can be chosen sequentially without replacement.

$s_0 + i_0 + r_0 = M$. The parameters $\alpha = \alpha(k)$ and $\gamma = \gamma(k)$ are considered to be time-varying because there are strong evidences that they change in the course of the epidemic, due to the reasons mentioned before. However, in some instances, the parameter $\alpha = \alpha(k)$ can be considered time-invariant, and determined if we know the half-life of the process with which infected individuals become removed, under the hypothesis that no contagion occurs. Considering N_r the half-life expressed in days, we must impose $(1-\alpha)^{N_r} = 1/2$, which allows us to determine

$$\alpha = 1 - 2^{-1/N_r} \tag{7}$$

that is, for a half-life of $N_r = 7$ days, $1/\alpha \approx 10$ is obtained, which seems to be quite reasonable, considering the observed data. By its turn, the parameter $\gamma = \gamma(k)$ indicates the rate at which the infection spreads over time and whenever it decreases, results on a gradual reduction of the number of infected people. A possible interpretation is that the parameter $\alpha(k)$ is a characteristic of the disease while $\gamma(k)$ results from the population behavior. For example, it depends on the social distancing adopted by the population in some time interval.

In epidemiology, there is a number that defines the secondary infections produced by one infected individual being introduced in a susceptible individuals group Hethcote (2000). This number (which in the present case depends on time) called basic reproduction number, denoted as R_0, in our time-varying SIR models is calculated as

$$\begin{aligned} R_0(k) &= \frac{\gamma(k)}{\alpha(k)} \\ &\geq \frac{\gamma(k)}{\alpha(k)} \left(\frac{s(k)}{M}\right)^{\nu} = R_\nu(k) \end{aligned} \tag{8}$$

Hence, from (5) it is clear that for values of $R_\nu > 1$, the infection spreads in the susceptible population, and on the contrary, whenever $R_\nu < 1$ the infection declines. The parameter R_0, an upper bound to R_ν, has a vital role in the study of epidemics and it helps us to observe how the epidemic is evolving in the population and approaches the end since $R_0 < 1$ implies that $R_\nu < 1$. Clearly, R_0 depends only on the model parameters (it does not depend on $s(k)$) and $R_0(k) \geq R_1(k) \geq R_2(k)$, for all $k \in \mathbb{N}$.

The proposed probabilistic SIR model has an intrinsic hypothesis that seems to be unrealistic. On each day, the average number of new infections is given by (2). To obtain this value, we assume that each individual in the population can meet any other, with equal probability. We believe that this simplifying hypothesis is no longer realistic when, for example, the population spreads over a large area with a non-uniform demographic density. The impact of this hypothesis, in face of reality, is difficult to measure. In fact, the possibility of all individuals meeting each other tends to increase the number of new infected, but not taking into account the eventual existence of high population densities, in some regions, acts in the opposite direction. Fortunately, as we will see later, this undesirable aspect can be mitigated if we consider time-varying models, as in (4)-(6). The same reasoning is valid for the classic SIR model for which, on each day, the average number of new infections is given by (1).

Finally, adding both sides of equations (4)-(6) it turns out that $s(k) + i(k) + r(k) = M$ for all $k \in \mathbb{N}$ which makes possible to express $r(k) = M - s(k) - i(k)$ and reduce the model to the following system of two nonlinear equations

$$s(k+1) = s(k) - \gamma(k) \left(\frac{s(k)}{M}\right)^{\nu} i(k) \tag{9}$$

$$i(k+1) = i(k) + \gamma(k) \left(\frac{s(k)}{M}\right)^{\nu} i(k) - \alpha(k) i(k) \tag{10}$$

with nonnegative initial conditions $s(0) = s_0$ and $i(0) = i_0$ such that $s_0 + i_0 \leq M$. As far as the time evolution is concerned the reduced order system (9)-(10) can be considered with no loss of generality and with the advantage that the trajectories evolve in the phase plane, a subset of $\mathbb{R}^2$ to be given in the next section.

3. STABILITY ANALYSIS

Dividing both equations (9) and (10) by the population size M, the one-to-one change of variables

$$(s(k)/M, i(k)/M, r(k)/M) \rightarrowtail (s(k), i(k), r(k))$$

shows that they must hold for the new variables as well. Hence, $M = 1$ can be fixed without loss of generality. Let us define the closed convex domain $\mathbb{D} \subset \mathbb{R}^2$ that plays a central role in the stability analysis of SIR models, that is

$$\mathbb{D} = \{(s, i) : s \geq 0, i \geq 0, s + i \leq 1\} \tag{11}$$

In addition, let us rewrite the previous model as $(s(k+1), i(k+1)) = Q_\nu \circ (s(k), i(k))$ where the nonlinear operator $Q_\nu : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$Q_\nu : \begin{pmatrix} s \\ i \end{pmatrix} \longmapsto \begin{pmatrix} s - \gamma s^\nu i \\ i + \gamma s^\nu i - \alpha i \end{pmatrix} \tag{12}$$

which exhibits the following important properties.

Lemma 1. Assume that $\alpha \in (0, 1)$. The set $\mathbb{D} \subset \mathbb{R}^2$ is an invariant set to the operator Q_ν, that is, $Q_\nu \circ \mathbb{D} \subseteq \mathbb{D}$ provided that:

(i) $\nu = 2$ and $\gamma \in (0, 3)$.
(ii) $\nu = 1$ and $\gamma \in (0, 1)$.

Proof: Denoting $(s_Q, i_Q) = Q_\nu \circ (s, i)$, considering $\alpha \in (0, 1)$ and $(s, i) \in \mathbb{D}$, it is immediate to verify that $i_Q \geq 0$ and $s_Q + i_Q = (s + i) - \alpha i \leq 1$ for both $\nu \in \{1, 2\}$. On the other hand, for any $\gamma > 0$, $\nu \in \{1, 2\}$, and $(s, i) \in \mathbb{D}$ it is seen that $s_Q = s - \gamma s^\nu i \geq s - \gamma s^\nu (1 - s) = g_\nu(s)$. Since $g_\nu(0) = 0$, two cases must be considered:

First, for $\nu = 2$, simple calculations put in evidence that $g'_\nu(s) \geq 1 - \gamma/3$. For $\gamma \in (0, 3)$ the function $g_\nu(s)$ is strictly increasing in the interval $s \in [0, 1]$. As a consequence $s_Q \geq 0$ for all $(s, i) \in \mathbb{D}$. Second, for $\nu = 1$, it can be verified that $g'_\nu(s) \geq 0$ for all $s \in [0, 1]$, provided that $|\gamma| < 1$. Hence, in the interval $\gamma \in (0, 1)$ we have that $s_Q \geq 0$ for all $(s, i) \in \mathbb{D}$. The proof is concluded. □

From Lemma 1, it is clear that for the proposed SIR model, the trajectories $(s(k), i(k)) \in \mathbb{D}$ for all $k \in \mathbb{N}$ provided that $(s_0, i_0) \in \mathbb{D}$, since its parameters are such that $\alpha(k) \in (0, 1)$ and $\gamma(k) \in (0, 2)$ for each $k \in \mathbb{N}$. For the classical SIR model the situation is more restrictive since this property is assured whenever the parameters satisfy $\alpha(k) \in (0, 1)$ and $\gamma(k) \in (0, 1)$. If this last condition is violated, it may occur that $s(k) < 0$ for some $k \in \mathbb{N}$, in which case, the meaning of the classic SIR model is lost.

Lemma 2. Assume that the conditions of Lemma 1 hold. For any initial condition $(s_0, i_0) \in \mathbb{D}$ the trajectory

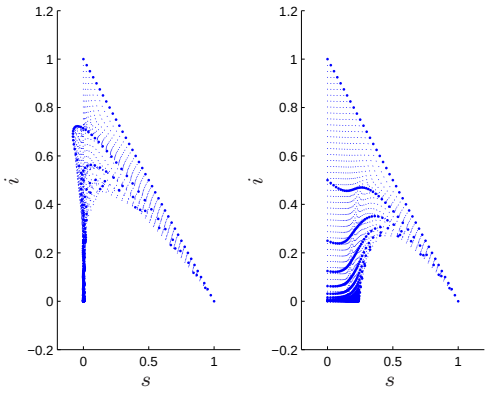

Fig. 1. Phase plane for $\gamma = 1.75$ and $\alpha = 0.50$

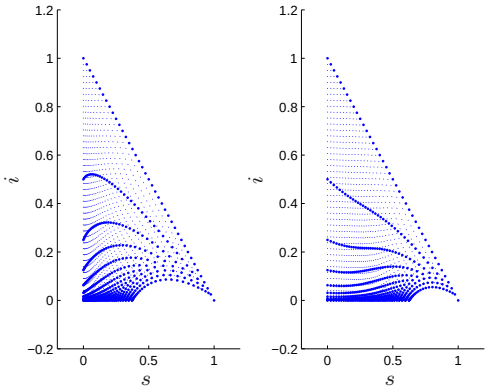

Fig. 2. Phase plane for $\gamma = 0.75$ and $\alpha = 0.50$

$(s(k), i(k))_{k\in\mathbb{N}} \in \mathbb{D}$ converges to an equilibrium point (s_*, i_*) satisfying the conditions $0 \leq s_* \leq s_0$ and $i_* = 0$.

Proof: Since any equilibrium point solves $(s_*, i_*) = Q_\nu \circ (s_*, i_*)$, it follows that $i_* = 0$. On the other hand, from Lemma 1, it has been established that $(s, i) \in \mathbb{D}$ provides $(s_Q, i_Q) \in \mathbb{D}$, which implies that the inequalities $0 \leq s_Q \leq s \leq 1$ hold as well. Consequently, the sequence $s(k),\ k \in \mathbb{N}$ converges to some $0 \leq s_* \leq 1$ because it is bounded below and non increasing putting in evidence that $s_* \leq s_Q$. Now consider the linear function $v(s, i) = (s - s_*) + i$ which is a valid Lyapunov function candidate for trajectories evolving in the region $(s, i) \in \mathbb{D}_* = \mathbb{D} \cap \{s \geq s_*\}$ of the phase plane. Simple algebraic manipulations yield

$$\begin{aligned} v(s_Q, i_Q) &= (s_Q - s_*) + i_Q \\ &= (s - s_*) + i - \alpha i \\ &\leq v(s, i) \end{aligned} \tag{13}$$

for all $(s, i) \in \mathbb{D}_*$, concluding thus the proof. □

For illustration we have drawn the phase plane, for some parameters, of the classic and proposed SIR models in discrete-time. Figure 1 has been determined for $\gamma = 1.75$ and $\alpha = 0.50$. On the left side part the phase plane of the classic SIR model ($\nu = 1$) is shown. It is clearly seen that, as expected, $\mathbb{D}$ is not an invariant set since the condition of Lemma 1 is violated. In this case, the stability property of Lemma 2 is no longer valid. For comparison, from the right side part of the same figure it is clear that $\mathbb{D}$ is an invariant set for the proposed SIR model ($\nu = 2$).

Figure 2 has been determined for $\gamma = 0.75$ and $\alpha = 0.50$. For both models, $\mathbb{D}$ is confirmed as an invariant set. For this particular choice of parameters, it is interesting to notice the similar behavior of both models but with the number of susceptibles at equilibrium s_* being bigger for the proposed model when compared to the one of the classic model. Under the same epidemic conditions, the proposed model appears to drawn a less severe situation as the classic model does. This important aspect needs factual confirmation.

4. PARAMETER IDENTIFICATION

In this section we consider the parameter identification problem where the goal is to determine the parameters $(\alpha(k), \gamma(k))$ and the initial condition from the available data. The World Health Organization as well as the Brazilian Ministry of Health daily reports the number of new diagnosed cases $n_m(k)$ and its accumulated sum $a_m(k)$. Notice that $n(k)$ becomes very different from $i(k)$ as the epidemic progresses. This is because infected people stop being infected when they transit to the removed class. The number of days of available data is denoted by N_m.

Defining the state space variable $z(k) = [s(k)\ i(k)\ a(k)]' \in \mathbb{R}^3$ with $a(k)$ being the accumulated number of new diagnosed cases, the coupling variable $w(k)$ and the output variable $y(k) = a(k)$, the unified version of the SIR model state space minimal realization can be written as

$$z(k+1) = A(k)z(k) + G(k)w(k),\ z(0) = z_0 \tag{14}$$

$$y(k) = Hz(k) \tag{15}$$

$$w(k) = f_\nu(z(k)) \tag{16}$$

where the indicated matrices are

$$A(k) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-\alpha(k) & 0 \\ 0 & 0 & 1 \end{bmatrix},\ G(k) = \begin{bmatrix} -\gamma(k) \\ \gamma(k) \\ \gamma(k) \end{bmatrix},\ H' = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

and the nonlinear function $f_\nu : \mathbb{R}^3 \to \mathbb{R}_+$ is

$$f_\nu(z(k)) = \left(\frac{s(k)}{M}\right)^\nu i(k) \tag{17}$$

Finally, it is important to keep in mind that the initial condition $z_0 = [s_0\ i_0\ a_0]' \in \mathbb{R}^3$ must be nonnegative and satisfy the constraint previously obtained $s_0 + i_0 \leq M$. This is indicated simply by $z_0 \in Z_0$. Additionally, the notation $(\alpha, \gamma) \in \Pi_\nu$ denotes the constraints $0 \leq \alpha \leq 1$ and $0 \leq \gamma \leq \nu$ whenever the classic SIR model ($\nu = 1$) or the proposed SIR model ($\nu = 2$) is concerned. Under these constraints Lemma 1 and Lemma 2 state that $\mathbb{D}$ is an invariant set and convergence towards the equilibrium point belonging to $\mathbb{D}$ is assured. In this section, two complementary situations are analysed. First, the parameters are supposed to be time-invariant which naturally imposes that they are constant during the entire time horizon of interest. Afterwards, the time-varying case with constant by parts parameters is treated.

4.1 Time-invariant Parameter Optimization

A well know procedure for parameter identification is adopted. It consists on the determination of the mean square error e_{ti} between the model and data through the

optimal solution of the nonlinear mathematical programming problem

$$\min_{z_0 \in Z_0, (\alpha,\gamma)\in\Pi_\nu} \frac{1}{2}\log_{10}\left(\frac{1}{N_m}\sum_{k=0}^{N_m-1}(y(k)-a_m(k))^2\right) \quad (18)$$

where $y(k)$ is the output provided by the model (14)-(17). It is worth mentioning that this problem is highly non-convex and, by consequence, only a local optimum is expected to be reached by the numerical procedure applied, see Mathworks (2005).

4.2 Time-varying Parameter Optimization

We assume that the time interval $[0, N_m)$ is subdivided into N sub-intervals $\{T_j\}_{j=1}^N$, without overlapping, such that $(\alpha(k), \gamma(k)) = (\alpha_j, \gamma_j)$ for all $k \in T_j$, and all $j = 1, \cdots N$. In other words, at each time interval the parameters to be determined remain constant. We have to determine e_{tv} from

$$\min_{z_0 \in Z_0, (\alpha_j,\gamma_j)\in\Pi_\nu} \frac{1}{2}\log_{10}\left(\frac{1}{N_m}\sum_{j=1}^{N}\sum_{k\in T_j}(y(k)-a_m(k))^2\right) \quad (19)$$

where as before, $y(k)$ is the output provided by the model (14)-(17). This problem is similar to (18). The only difference between them is the number of variables to handle. Moreover, since the constraints $(\alpha_j, \gamma_j) \in \Pi_\nu$ for $j = 1, \cdots, N$ are decoupled, at the optimal solution, the minimum cost naturally satisfies $e_{tv} \leq e_{ti}$ because any solution to (18) is feasible to (19). This aspect will be confirmed numerically by the examples solved.

4.3 Sequential Forward Optimization

In the time-varying parameter optimization context, whenever new measurements are treated during the outbreak evolution, the whole time evolution of all parameters can be modified. To preserve optimal past values, a strategy inspired on receding horizon seems to be well adapted, see Bemporad et al. (2002) for details. In other words, the future of the outbreak evolution can not modify the values of parameters in the past and present. This imposes causality to the parameter identification procedure. It can be stated by considering again that the parameters $(\alpha(k), \gamma(k))$ are constant by parts. At an arbitrary time sub-interval T_j for some $j = 1, \cdots, N$, we need to solve

$$e_j^2 = \min_{(\alpha_j,\gamma_j)\in\Pi_\nu} \sum_{k\in T_j}\Big(y(k)-a_m(k)\Big)^2 \quad (20)$$

where $y(k)$ is the output provided by the model (14)-(17) starting from the initial condition $z_0 = [(M - a_m(0))\ a_m(0)\ a_m(0)]' \in \mathbb{R}^3$. Proceeding in this way it is possible to determine the parameters $(\alpha(k), \gamma(k))$ for all $k \in [0, N_m)$. Finally, the mean square error

$$e_{so} = \frac{1}{2}\log_{10}\left(\frac{1}{N_m}\sum_{j=1}^{N} e_j^2\right) \quad (21)$$

between the sequence $a(k)$ determined through the proposed time-varying model and the corresponding measured values $a_m(k)$, actually observed, gives a measure of the model adherence to reality. Compared to the previous strategy this one is simpler but sub-optimal implying that $e_{tv} \leq e_{so}$. However, in general, it has been verified that $e_{so} < e_{ti}$ whenever the time sub-intervals $T_j,\ j = 1, \cdots, N$ are appropriately chosen.

Finally, from the previous results it is important to mention that, with precaution, since estimation errors can be expressive, epidemics short-term evolution can be estimated by keeping the parameters constant, and equal to the values identified during the last time sub-interval, that is $(\alpha(k), \gamma(k)) = (\alpha_N, \gamma_N)$ for all $k \geq N_m$, see Geromel et al. (2002).

5. SIMULATION AND VALIDATION

In this section two outbreak evolutions are analysed in detail. First, the outbreak in Brazil, which is until now in franc expansion, is considered. Afterwards, the outbreak in Italy, which already reached the end, is handled through the same proposed model in order to put in evidence the adherence to data and precision. It is important to mention that all parameters identification problems have been solved with MatLab Version R14, see Mathworks (2005) for details and, to avoid undesirable numerical singularity on the determination of $R_0(k)$ from (8), we have included the constraint $\alpha(k) \geq 1/10$ for all $k \in [0, N_m)$ which is not effective since the minimum mean square errors remain approximately unchanged.

5.1 Outbreak in Brazil

Since the first reported Brazilian case up to the day this article was written, 146 days have passed, according to official data [5] provided in MS (2020) and WHO (2020). The identification error produced by each method presented before with $N_m = 146$ [day], $N = 9$ and $T_1 = [0, 28)$, $T_2 = [28, 42)$, $T_3 = [42, 56)$, $T_4 = [56, 70)$, $T_5 = [70, 84)$, $T_6 = [84, 98)$, $T_7 = [98, 112)$, $T_8 = [112, 126)$, $T_9 = [126, 146)$ was $e_{ti} = 4.75$, $e_{tv} = 3.72$, and $e_{so} = 3.72$. The first interesting aspect is $e_{tv} \approx e_{so} < e_{ti}$ which means that the sub-optimality imposed by the sequential forward optimization is negligible. However, the identified parameters trajectories associated to the minimum errors e_{tv} and e_{so} are very different. Indeed, Figure 3 shows the time evolution of $R_0(k)$, becoming clear that at the beginning they are different but they approach to the same value (≈ 1.01) as the infection spreads in the population. Due to the fact that, until this moment, the number of susceptible individuals is very close to the population ($\approx 99\%$), it can be verified that $R_0(k) \approx R_\nu(k)$ for all $k \in [0, N_m)$ and $\nu \in \{1, 2\}$, see (9). As mentioned before, we have considered $M = 210$ million inhabitants, IBGE (2020).

With the parameters $(\alpha(k), \gamma(k))$ obtained from the *Sequential Forward Optimization* procedure, we have compared the long-term behavior of the classic ($\nu = 1$) and the proposed variant ($\nu = 2$) of the SIR model. As already mentioned, until the $126^{\underline{th}}$ epidemic day, the outcome of both models are practically identical. Figure 4 shows the outbreak evolution one more year from now, approximately. On the top the estimated number of new daily cases is shown. While the classic SIR model estimates a

[5] Data from both cited sources are slightly different. We have used those provided by WHO.

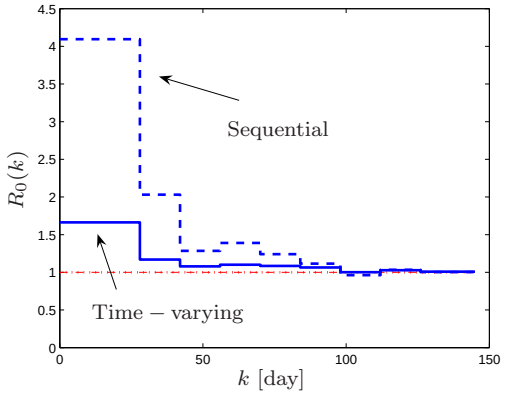

Fig. 3. Basic reproduction number evolution in Brazil

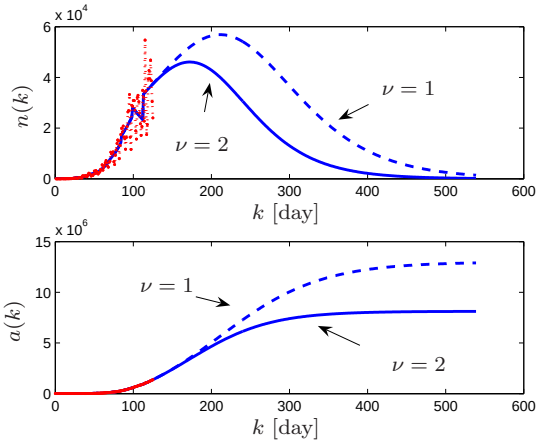

Fig. 4. Outbreak long-term time evolution in Brazil

maximum of about sixty thousand, for the proposed model this number reduces to forty thousand. As a consequence, on the bottom of Figure 4, for both models, the accumulate sum of daily new cases is shown. The reduction previewed by the proposed SIR model in comparison to the classical one is expressive. In other words, from now to the end of the outbreak evolution, the difference between the outcomes of both models is significant, even though their behaviour until now are very similar. This claim is supported by official data plotted in red marks "·" in both parts of the aforementioned figure. However, it is important to make clear that these situations may not be confirmed due to parameter changing that may occur in the future and obviously can not be taken into account in the present. This aspect will be discussed in the sequel.

5.2 Outbreak in Italy

The outbreak in Italy reached the end after 124 days. Data for the entire evolution of the epidemic is available WHO (2020) and, consequently, all stages can be taken into account for parameter identification. Hence, we have considered $N_m = 124$ [day], $N = 5$ and the time subintervals $T_1 = [0, 42)$, $T_2 = [42, 49)$, $T_3 = [49, 77)$, $T_4 = [77, 98)$, $T_5 = [98, 124)$. The values of the parameter identification error were $e_{ti} = 4.44$, $e_{tv} = 2.61$, and $e_{so} = 2.59$. Some aspects must be put in evidence. First, the large value of $e_{ti} \gg e_{tv} \approx e_{so}$ indicates that the model with time-invariant parameters is inappropriate because

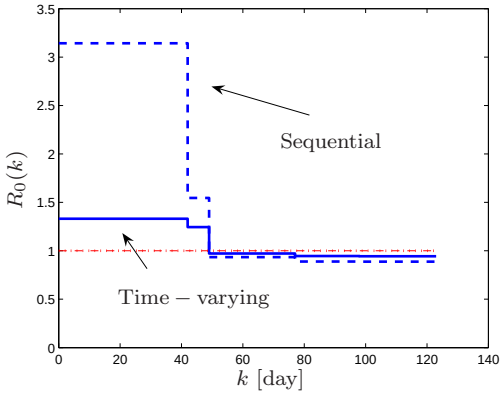

Fig. 5. Basic reproduction number evolution in Italy

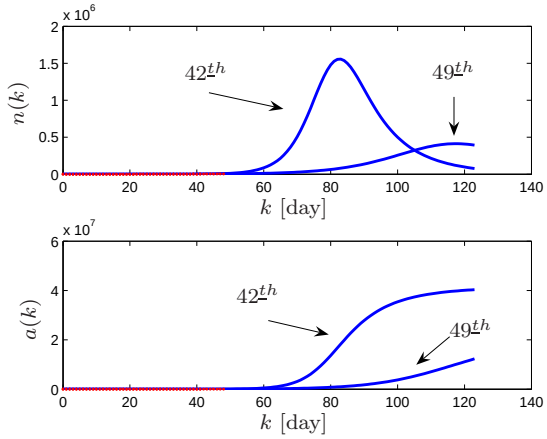

Fig. 6. Outbreak long-term time evolution in Italy

the minimum error is very expressive. Second, the total of accumulated cases is very small (less that 0.4%) when compared to the population of $M = 60$ million inhabitants which naturally means that both models with $\nu = 1$ or $\nu = 2$ provide virtually the same solutions. For this reason we have adopted, from now on, the proposed SIR model with time-varying parameters determined by the *Sequential Forward Optimization* procedure.

Figure 5 shows the basic reproduction number provided by the solution of problems (19) and (20), respectively. The optimal solution of (20) corresponding to the first time interval, namely $T_1 = [0, 42)$ satisfies exactly the lower bound $\alpha_1 = 1/10$ indicating that in this time subinterval the value of the basic reproduction number may be very high. This fact occurs in the first time interval only. However, if this lower bound is removed, it has been verified that the minimum error e_{so} provided by the optimal solution of problem (20) remains practically the same. Figure 5 makes clear that, as the outbreak progresses, the basic reproduction numbers approach each other.

We now move our attention to the long-term behavior prediction using this epidemic model. It is well known that long-term behavior prediction is very hard to perform in the context of dynamic systems with time-varying parameters. Indeed, parameters identification based on past and present data may define a precise model for a

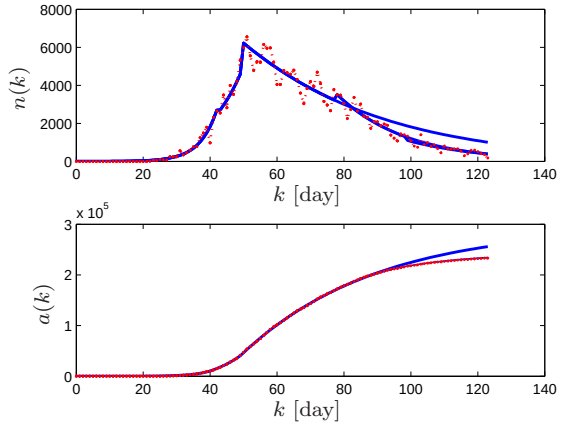

Fig. 7. Outbreak time evolution in Italy

certain time interval, but it may become poor in terms of precision in the future face the possibility of parameter changing. To illustrate this claim we have simulated the proposed model in several situations.

First, supposing that we are at the $42^{\underline{th}}$ epidemic day, with the available data, we have adopted the *Sequential Forward Optimization* procedure to estimate the accumulated number of new cases for all epidemic days (label $42^{\underline{th}}$). In a second run, we did the same but considering the available data until the $49^{\underline{th}}$ epidemic day (label $49^{\underline{th}}$). Figure 6 shows the number of new cases (on the top) and the accumulated number of new cases (on the bottom), respectively, and data with red marks "·". Figure 7 has been obtained exactly in the same way assuming that we are at the $77^{\underline{th}}$, $98^{\underline{th}}$ and $124^{\underline{th}}$ epidemic day, respectively. Until the $49^{\underline{th}}$ day the outbreak proceeds in *acceleration phase*, the model is very precise but long-term prediction is very poor and with clear lack of precision. It suffices to compare the predicted maximum number of daily new cases (about $500,000$ occurring at the $118^{\underline{th}}$ day) with the true value (about $7,000$ occurred at the $49^{\underline{th}}$ day).

The same difficulty is not observed when the available data correspond to the $77^{\underline{th}}$ and the $98^{\underline{th}}$ epidemic days because the social distancing and other measures adopted in the country moved the outbreak behaviour to a *retraction phase*. The model fits very well to data and prediction is precise as well. This claim is clearly supported by verifying the red marks "·", and solid curves in Figure 7. The ones corresponding to data available until the $98^{\underline{th}}$ epidemic day and all data available until the end of the epidemic are practically identical. This leads to the conclusion that more research effort must be done towards the development of more accurate prediction models depending on time-varying parameters. Another possibility, successfully adopted in Giordano et al. (2020), is to consider scenarios defined by parameters leading to prescribed $R_0(k)$ and evaluating by the model the impact of each scenario on the epidemic evolution. To accomplish this goal, more precise models are essential.

6. CONCLUSION

The dynamic model proposed in this paper has good adherence to reality, however, due to the presence of time-varying parameters, it does not allow reliable long-term prediction of the epidemic evolution. After all, if the parameters change in the future, there is no way to estimate them from observations of the past and present. Fortunately, the reported results seem to indicate that this fact becomes less important as new data is processed, which makes possible short-term prediction. Hence, the impact of well-defined scenarios on the epidemic evolution can be done with accuracy. For long-term prediction time-varying parameters modeling seems to be essential to increase accuracy. It is important to mention that in this paper we proposed an alternative probabilistic dynamic model of SIR class that, in principle, can be generalized to obtain continuous and discrete-time models like SEIR, SEIRS, SIRS, SEI, SEIS, SI, SIS, among others. Finally, we would like to emphasize once again that its practical viability needs factual confirmation.

REFERENCES

Anastassopoulou, C., Russo, L., Tsakris, A., and Sietto, C. (2020). Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PLOS ONE*. Doi: 10.1371/journal.pone.0230405.

Anderson, R.M. (1987). *The Population Dynamics of Infectious Diseases: Theory and Applications*. Springer, London.

Bailey, N.T.J. (1975). *The mathematical theory of infectious diseases and its applications*. Hafner Press, 2nd ed., New York.

Bemporad, A., Borrelli, F., and Morari, M. (2002). Model predictive control based on linear programming - the explicit solution. *IEEE Transac. on Automatic Control*, 47, 1974–1985.

Bertozzi, A.L., Franco, E., Mohler, G., Short, M.B., and Sledge, D. (2020). The challenges of modeling and forecasting the spread of covid-19. *Proceedings of the National Academy of Sciences*, 29, 16732–16738.

Brauer, F. and Castillo-Chavez, C. (2012). *Mathematical Models in Population Biology and Epidemiology*. Springer, New York.

Geromel, J.C., de Oliveira, M.C., and Bernussou, J. (2002). Robust filtering of discrete-time linear systems with parameter dependent lyapunov functions. *SIAM Journal on Control and Optimization*, 41, 700–711.

Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Filippo, A.D., Matteo, A.D., and Colaneri, M. (2020). Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, 26, 855–860.

Hethcote, H.W. (2000). The mathematics of infectious diseases. *SIAM Review*, 42, 599–653.

IBGE (2020). *Area and Population of Brazil (in Portuguese)*. Instituto Brasileiro de Geografia e Estatística.

Kermack, W.O. and McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London - A*, 115, 700–721.

Mathworks, I. (2005). *MATLAB, Version R14 (7.1.0.246)*. Mathworks, Inc.

MS (2020). *Painel Coronavirus*. Brazilian Ministry of Health.

WHO (2020). *Coronavirus disease (COVID-2019) situation reports*. World Health Organization.